Probabilistic Case Detection

Jeremy Espino MD

RODS Laboratory

Center for the Advanced Study of Informatics in Public Health

CDC Center of Excellence in Public Health Informatics

University of Pittsburgh

Outline

- Problem
- Description of the case detection system
- Demonstration
- Evaluation

Problem

Existing automated disease reporting systems

are incomplete

- ELR can not identify probable or suspect cases that use clinical case criteria
- Syndromic surveillance is non specific

are delayed

- ELR relies on final reports

and cannot adjust their output to the prevalence of the disease.

Solution: Probabilistic Case Detection

- Use data from the entire patient visit including preliminary data
- Use structured and coded data; extracted when necessary from free text
- Utilize a Bayesian expert system which integrates disease prevalence (prior probability) and findings to provide definitive as well as probable diagnoses

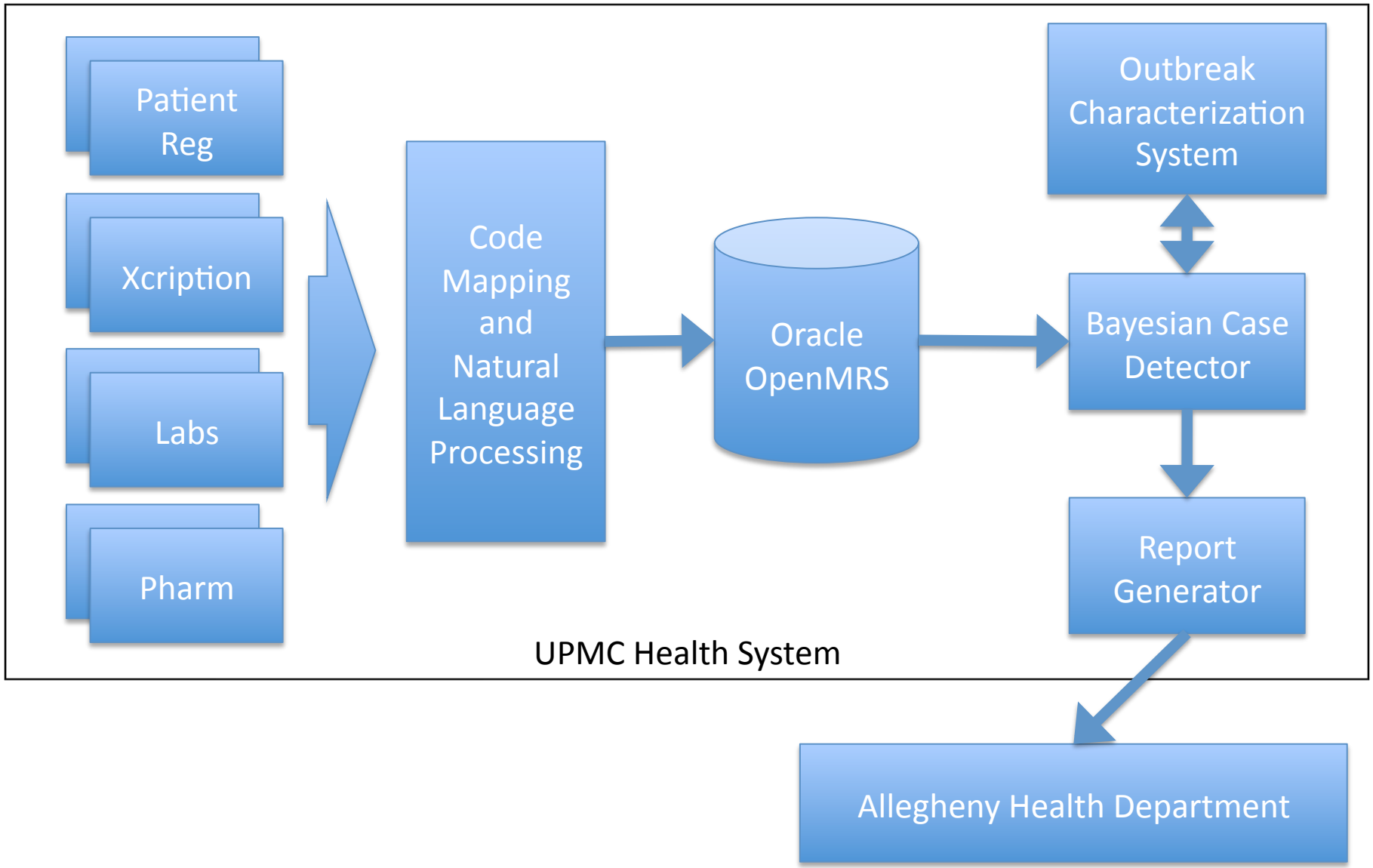
Bayes Theorem

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}.$$

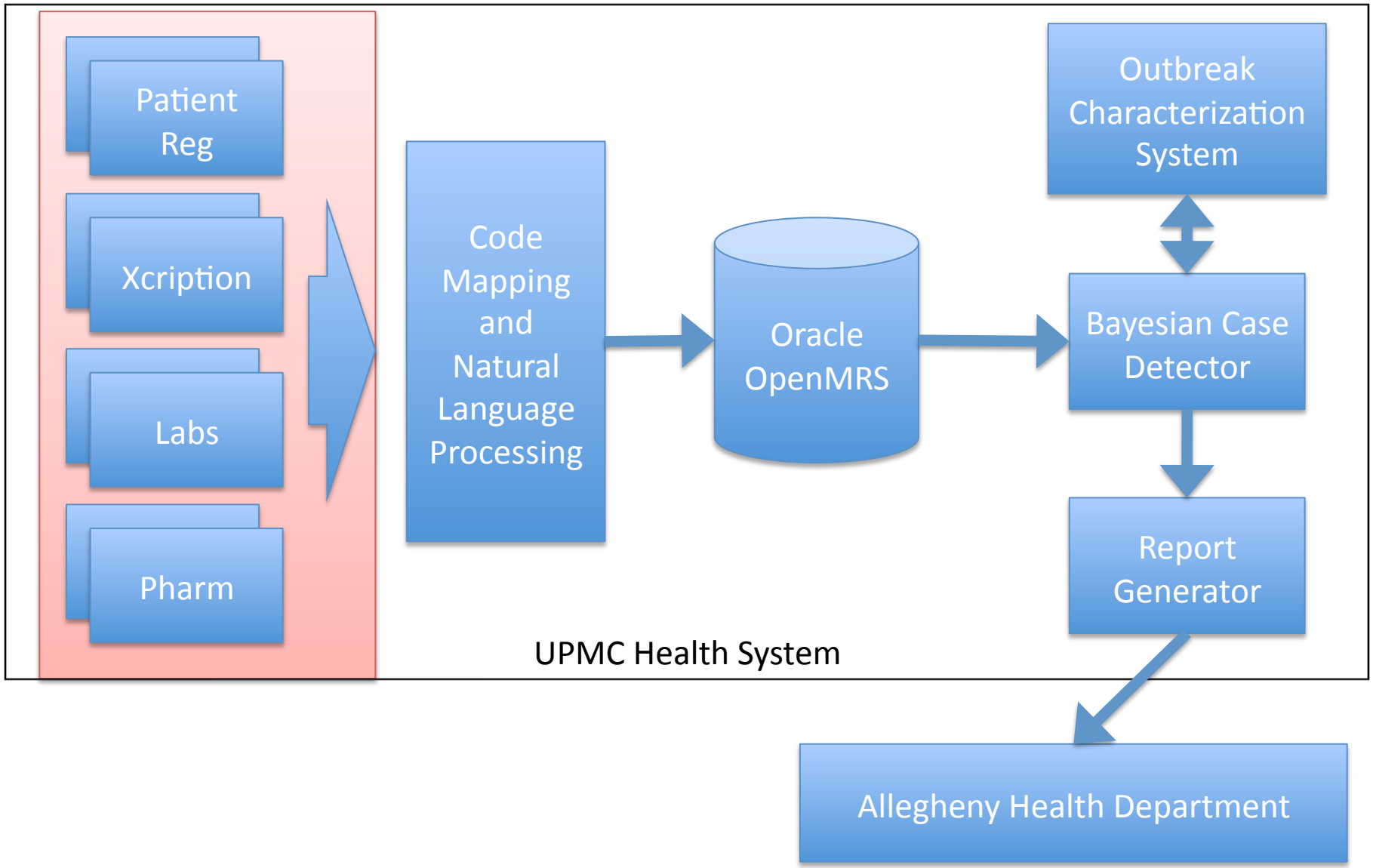
- P(A|B) the posterior probability of a disease
- P(A) is the prior probability of the disease (i.e., the prevalence)
- P(B|A) is the conditional probability of some finding given the disease.
- P(B) the marginal likelihood of the finding

System Description

Probabilistic Case Detection System



Probabilistic Case Detection System



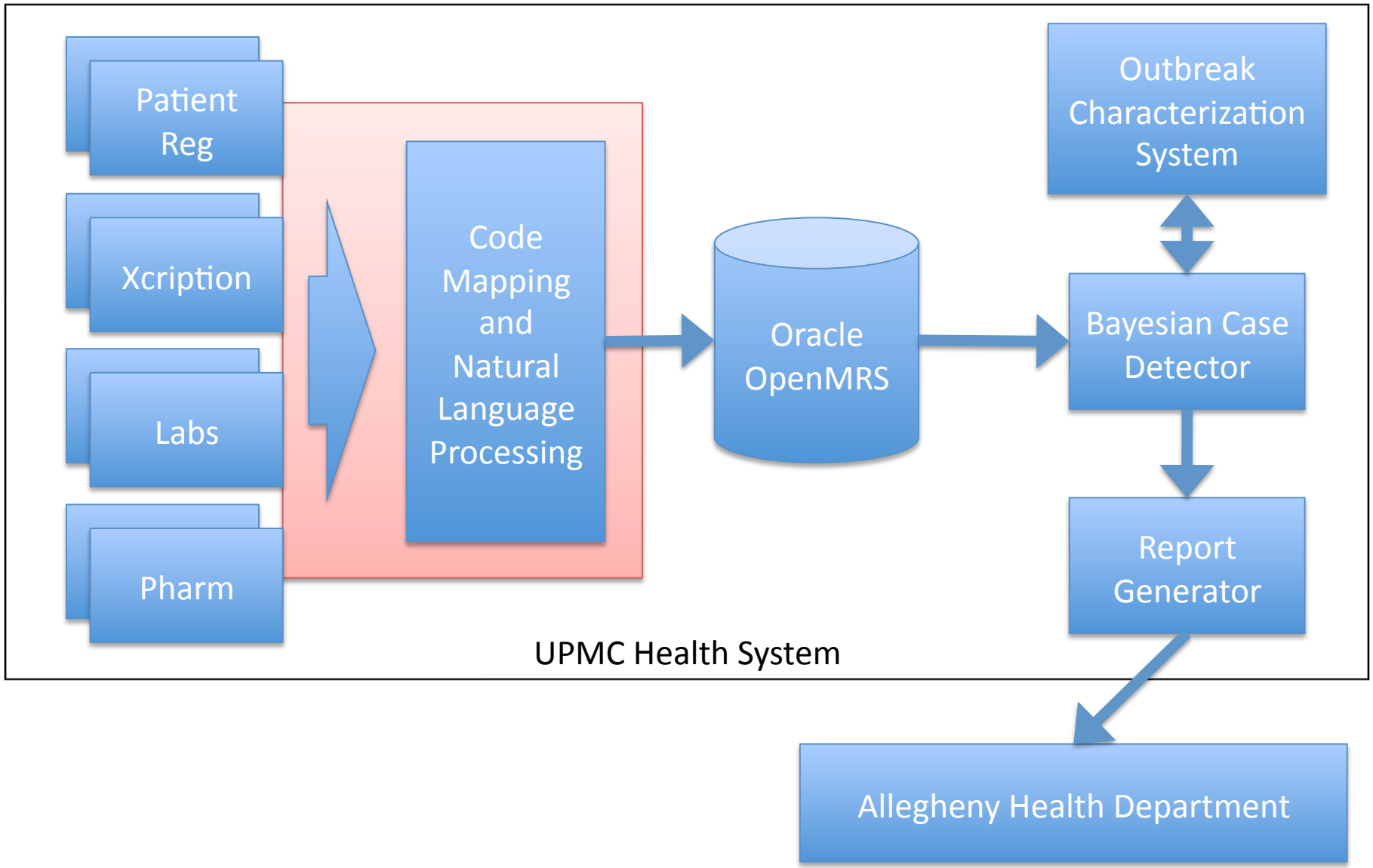
Data Sources

- Our case detection system operates within the healthcare system firewalls
- We receive HL7 messages from the healthcare system's enterprise router
- We transform these messages into OpenMRS Observation records using Groovy scripts and store them in an Oracle database

HL7 Message Volume

- 13 hospitals
- 84,000 patient registrations/day
- 5000 microbiology results/day
- 35,000 chemistry results/day
- 5,000 text reports/day

Probabilistic Case Detection System



Structured / Coded Data

- **DATA_TYPE (DATA_TYPE):** ADT_A04 (ADT_A04)
  - **PATIENT_CLASS (PATIENT_CLASS):** P
  - **PATIENT_LOCATION (PATIENT_LOCATION):** OTSN
  - **ATTENDING_DOCTOR_ID (ATTENDING_DOCTOR_ID):** 07857
  - **ATTENDING_DOCTOR_NAME (ATTENDING_DOCTOR_NAME):** [redacted]
  - **REFERING_DOCTOR_ID (REFERING_DOCTOR_ID):** 06152
  - **REFERING_DOCTOR_NAME (REFERING_DOCTOR_NAME):** [redacted]
  - **CONSULTING_DOCTOR_ID (CONSULTING_DOCTOR_ID):** 02989
  - **CONSULTING_DOCTOR_NAME (CONSULTING_DOCTOR_NAME):** [redacted]
  - **ADMITTING_DOCTOR_ID (ADMITTING_DOCTOR_ID):** 04947
  - **ADMITTING_DOCTOR_NAME (ADMITTING_DOCTOR_NAME):** [redacted]
  - **HOSPITAL_SERVICE (HOSPITAL_SERVICE):** OTS
  - **FINANCIAL_CLASS (FINANCIAL_CLASS):** X
  - **HOSPITAL_CODE (HOSPITAL_CODE):** WW
  - **EMPLOYER_ADDR_STREET (EMPLOYER_ADDR_STREET):** 5207 [redacted]
  - **EMPLOYER_ADDR_CITY (EMPLOYER_ADDR_CITY):** PA
  - **EMPLOYER_ADDR_STATE (EMPLOYER_ADDR_STATE):** 16673
  - **RAW_HL7 (RAW_HL7):** MSH|^~\&|SMS|WW|RS ROUTER|RCC|201004231737||ADT^A04|000002147483647|P|2.2| PID|| [redacted] ...
- **DATA_TYPE (DATA_TYPE):** LAB_MICRO (LAB_MICRO)
  - **RELEVANT_CLINICAL_INFORMATION (RELEVANT_CLINICAL_INFORMATION):** URNC14 (URNC14)
  - **SPECIMENT_RECEIVED_DATE (SPECIMENT_RECEIVED_DATE):** [redacted] 2010
  - **REPORTED_DATE (REPORTED_DATE):** [redacted] 2010
  - **PATIENT_LOCATION (PATIENT_LOCATION):** 3NB^3[redacted]
  - **OBS_GROUP (OBS_GROUP):** Organism
    - **Organism (ORG):** PROTEUS MIRABILIS (PRMI)
    - **METHOD (MTYP):** Kirby Bauer Urines (KBUR)
    - **Ampicillin/Sulbactam (AMSUL):** Sensitive (SS)
    - **Cephalothin (CEPH):** Sensitive (SS)
    - **Cefuroxime (CXM):** Sensitive (SS)
    - **Ceftriaxone (CRO):** Sensitive (SS)

Natural Language Processing

- Extract UMLS concepts from free text
  - Medlee (Columbia University)
  - Onyx (University of Pittsburgh)
  - Topaz (University of Pittsburgh)
- Identify medical terms using a large dictionary along with variants
- Identify negated terms
- Determine if the extracted concepts are chronic/acute and current/historical

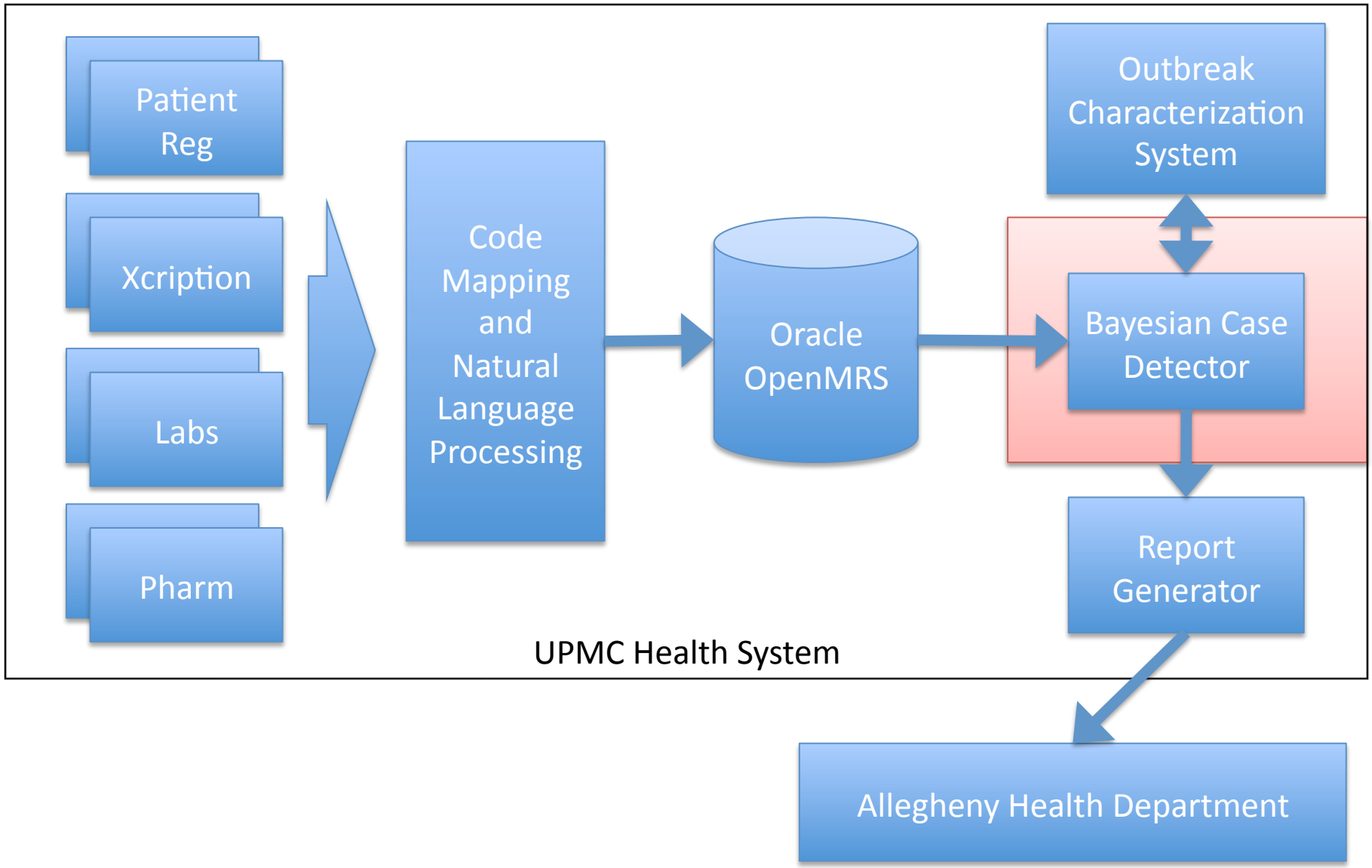
Example: Free Text

ASSESSMENT/PLAN:
1. Cardiovascular/permanent pacemaker battery reaching end of life/status postpermanent pacemaker Guidant implant on **DATE[Jun 02], 2009, following a syncopalevent, with insertion of right arterial and coronary sinus lead and withbattery change to a CRT-P generator for end of life on **DATE[Jun 02], 2009/ ischemiccardiomyopathy/CAD/MI/CHF/diastolic dysfunction - denies chest pain; mildshortness of breath.
2. Pulmonary/obstructive sleep apnea - does not use CPAP or BiPAP; insurancedoes not cover home oxygen therapy; denies respiratory distress.
3. Hypertension - stable; no nosebleeds or headaches.
4. Hyperlipidemia - stable; no evidence of myopathy or hepatitis.
5. Prostate carcinoma/status post radiation therapy - currently in remission;no evidence of recurrence.
6. GI/diverticulitis - denies any GI symptomatology.
7. Obesity/hypoalbuminemia/catabolic protein malnutrition - encourage good p.o.intake; consider calorie count; consider nutritional consult; monitornutritional values; most recent albumin and total protein value drawn on **DATE[Jun]09, 2009, equals 3.0 and 5.7; weight reduction necessary in the future.
8. Anemia - most recent hemoglobin and hematocrit level drawn on **DATE[Jun 10], 2009,equals 10.9 and 31.9; no evidence of GI bleed or retroperitoneal hemorrhage.
9. Thrombocytopenia - most recent platelet count drawn on **DATE[Jun 10], 2009, equalshemorrhage.10. Deconditioning - continue with physical therapy, occupational therapy,PM\T\R, recreational therapy, and good nutritional support.

Example: Extracted Concepts

- **NLP_PROCESSOR (NLP_PROCESSOR):** MEDLEE (MEDLEE)
- **OBS_GROUP (OBS_GROUP):** FINDING
  - **FINDING (FINDING):** Anemia (C0002871)
  - **SECTION_NAME (SECTION_NAME):** report assessment item
  - **SNIPPET (SNIPPET):** Anemia most recent hemoglobin hematocrit level drawn
  - **STATE (STATE):** T (T)
  - **TYPE (TYPE):** SYMPTOM (SYMPTOM)
- **OBS_GROUP (OBS_GROUP):** FINDING
  - **FINDING (FINDING):** Chest Pain (C0008031)
  - **SECTION_NAME (SECTION_NAME):** report assessment item
  - **SNIPPET (SNIPPET):** Cardiovascular / permanent pacemaker battery reaching e
  - **STATE (STATE):** F (F)
  - **TYPE (TYPE):** SYMPTOM (SYMPTOM)
- **OBS_GROUP (OBS_GROUP):** FINDING
  - **FINDING (FINDING):** Diverticulitis (C0012813)
  - **SECTION_NAME (SECTION_NAME):** report assessment item
  - **SNIPPET (SNIPPET):** GI / diverticulitis denies GI symptomatology
  - **STATE (STATE):** T (T)
  - **TYPE (TYPE):** SYMPTOM (SYMPTOM)
- **OBS_GROUP (OBS_GROUP):** FINDING
  - **FINDING (FINDING):** Epistaxis (C0014591)
  - **SECTION_NAME (SECTION_NAME):** report assessment item
  - **SNIPPET (SNIPPET):** no nosebleeds headaches
  - **STATE (STATE):** F (F)
  - **TYPE (TYPE):** SYMPTOM (SYMPTOM)
- **OBS_GROUP (OBS_GROUP):** FINDING
  - **FINDING (FINDING):** Gastrointestinal Hemorrhage (C0017181)
  - **SECTION_NAME (SECTION_NAME):** report assessment item
  - **SNIPPET (SNIPPET):** no evidence of GI bleed or retroperitoneal hemorrhage
  - **STATE (STATE):** F (F)

Probabilistic Case Detection System



Measles (Rubeola)

1990 Case Definition

Clinical case definition

An illness characterized by all of the following clinical features:

- a generalized rash lasting greater than or equal to 3 days
- a temperature greater than or equal to 38.3°C (101°F)
- cough, or coryza, or conjunctivitis

Laboratory criteria for diagnosis

- Isolation of measles virus from a clinical specimen, or
- Significant rise in measles antibody level by any standard serologic assay, or
- Positive serologic test for measles IgM antibody

Case classification

Suspect: any rash illness with fever

Probable: meets the clinical case definition, has no or noncontributory serologic or virologic testing, and is not epidemiologically linked to a probable or confirmed case

Confirmed: a case that is laboratory confirmed or that meets the clinical case definition and is epidemiologically linked to a confirmed or probable case. A laboratory-confirmed case does not need to meet the clinical case definition.

Comment

Two probable cases that are epidemiologically linked would be considered confirmed, even in the absence of laboratory confirmation. Only confirmed cases should be reported to the NNDSS.

Rubeola Criteria

- Clinical Criteria – all of the following signs and symptoms
  - generalized rash >= 3 days
  - Temperature >= 101F
  - Cough, coryza, or conjunctivitis
- Lab Criteria - one of the following
  - Isolation of measles virus from clinical specimen
  - Significant rise in measles antibody level
  - Positive serologic test for measles IgM antibody

Rubeola Case Classification

- Suspect – any rash illness with fever
- Probable
  - Meets clinical case definition
  - Does not meet lab criteria
- Confirmed
  - Meets case criteria
  - Meets lab criteria
  - Epidemiologically linked to confirmed or probable case

  OR
  - Two probable cases epidemiologically linked are considered confirmed

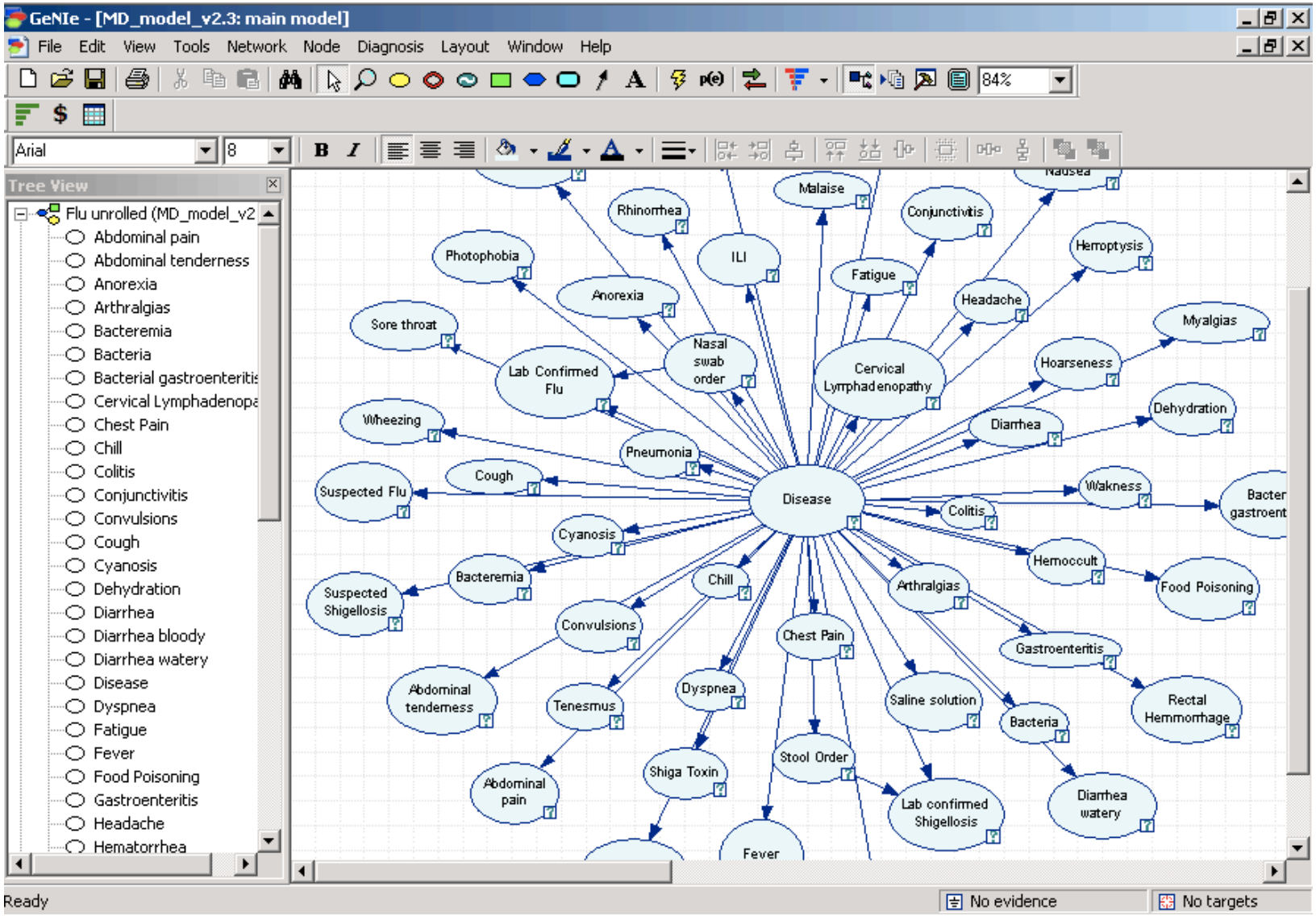
Probabilistic Case Definition Recipe

- Required Ingredients
  - Disease prior (prevalence)
  - Disease and lab criteria
  - Conditional probabilities of criteria given disease
- Optional Ingredients
  - Performance characteristics of tests
  - Performance characteristics of NLP extraction
- The required ingredients can be obtained from case definitions, literature, an expert and/or data

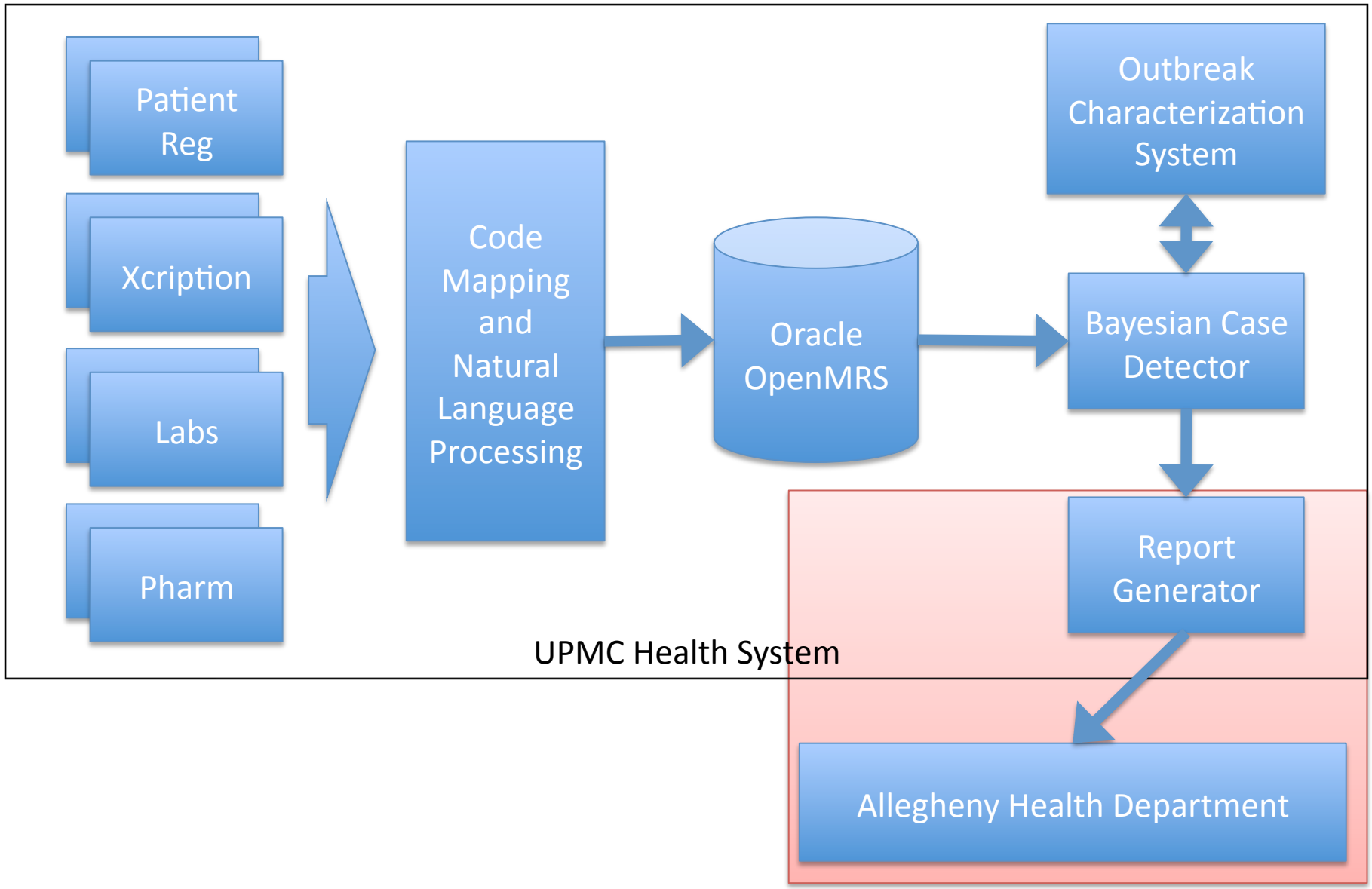
Probabilistic Case Definition

- Use the ingredients to "cook up" a Bayesian network
- The Bayesian network is a representation of the conditional dependencies between a set of random variables
- GENIE – Bayesian network modeling tool created at the University of Pittsburgh

GENIE - Bayesian Network Editor



Probabilistic Case Detection System

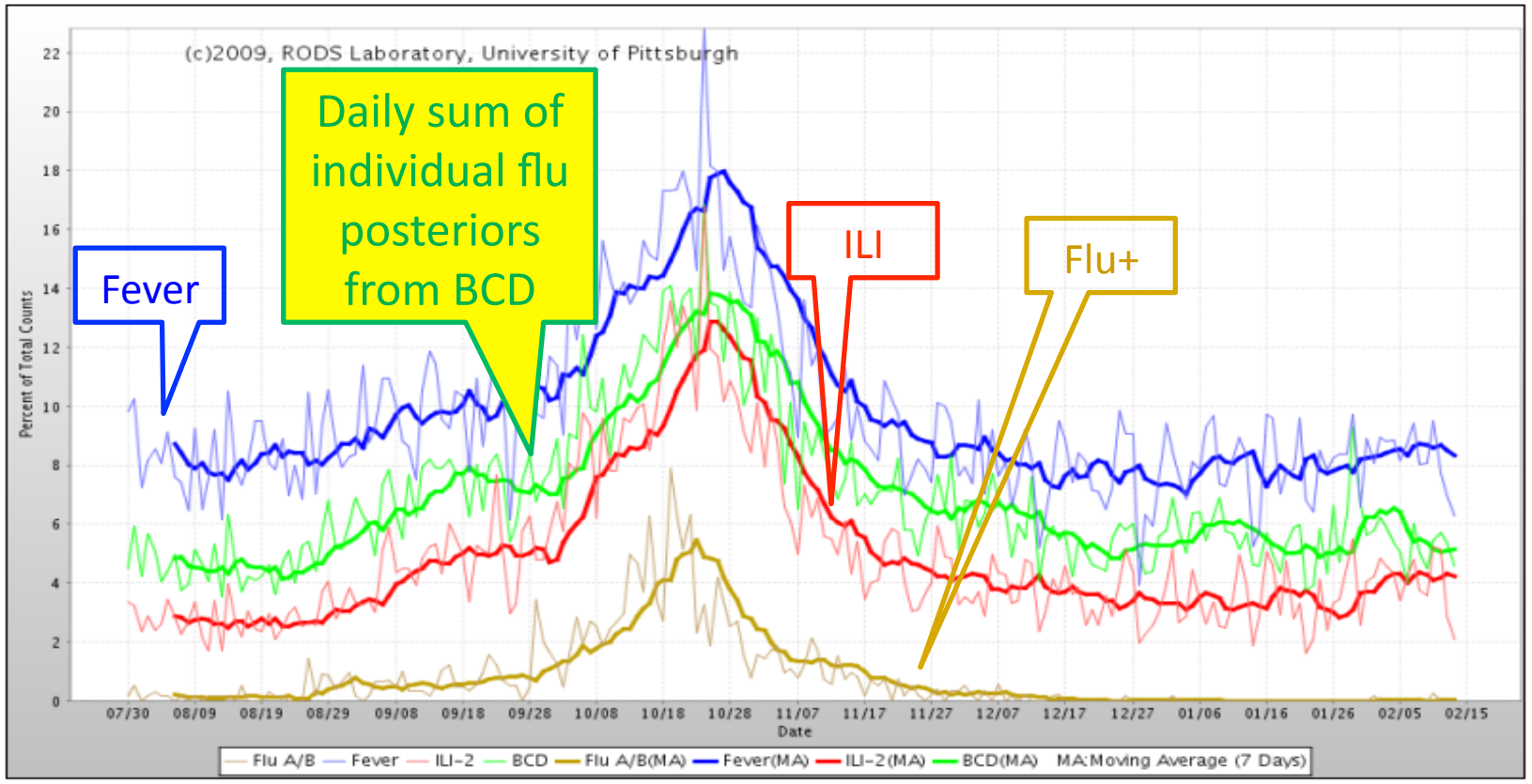


Daily Email Report to ACHD

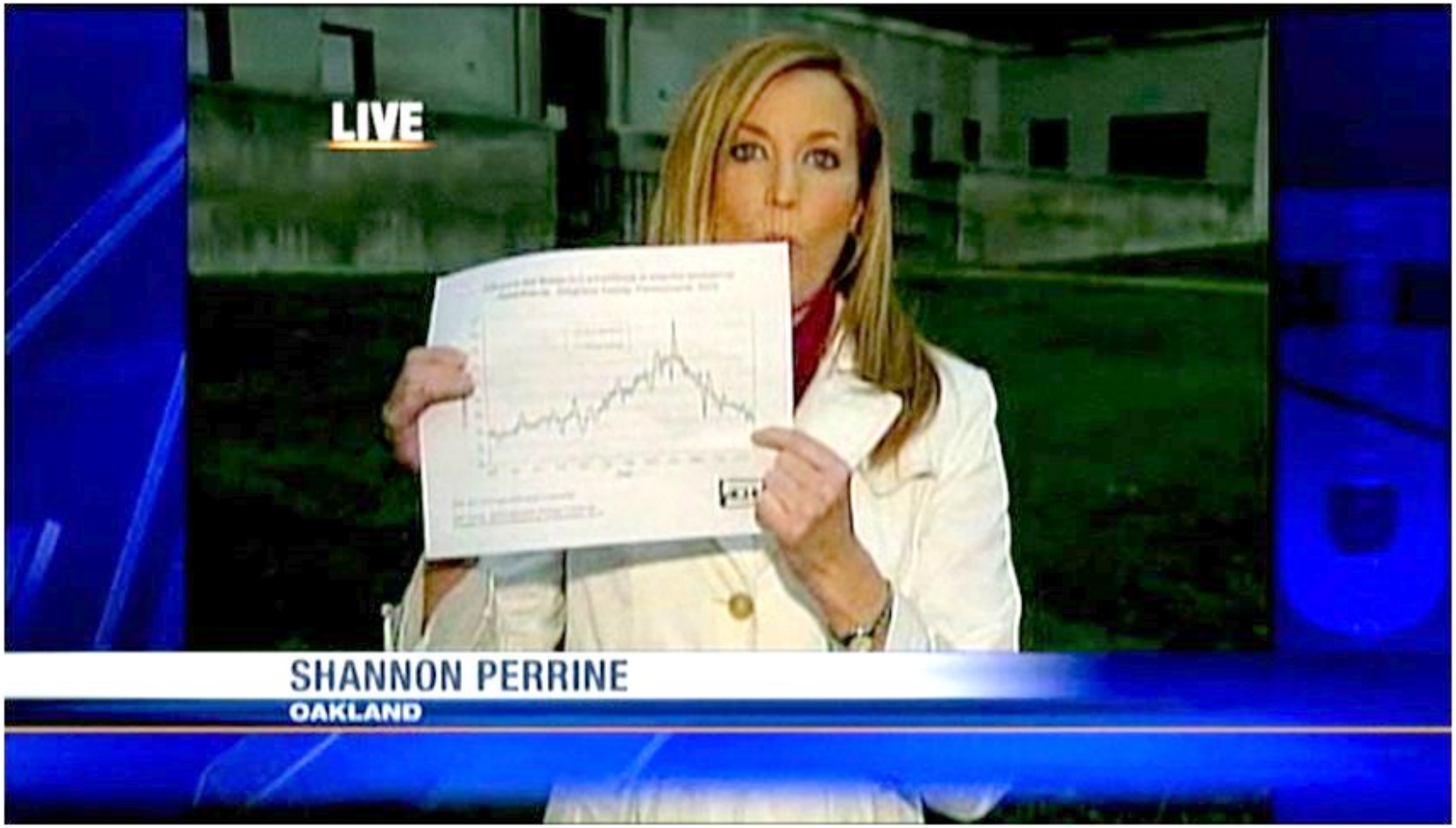
Daily ILI and fever reports - Message (Plain Text)

Size: 125 KB
Image dimensions: 1000 x 600

Message | summary_chart_20100214.png (133 KB) | summary_chart2_20100214.png (125 KB) | summary_20100214.csv (357 B)
fever_by_hosp_20100214.csv (1 KB) | fever_by_zip_20100214.csv (9 KB) | ili_by_hosp_20100214.csv (1 KB)
ili_by_zip_20100214.csv (8 KB) | ili2_by_hosp_20100214.csv (1 KB) | ili2_by_zip_20100214.csv (8 KB)



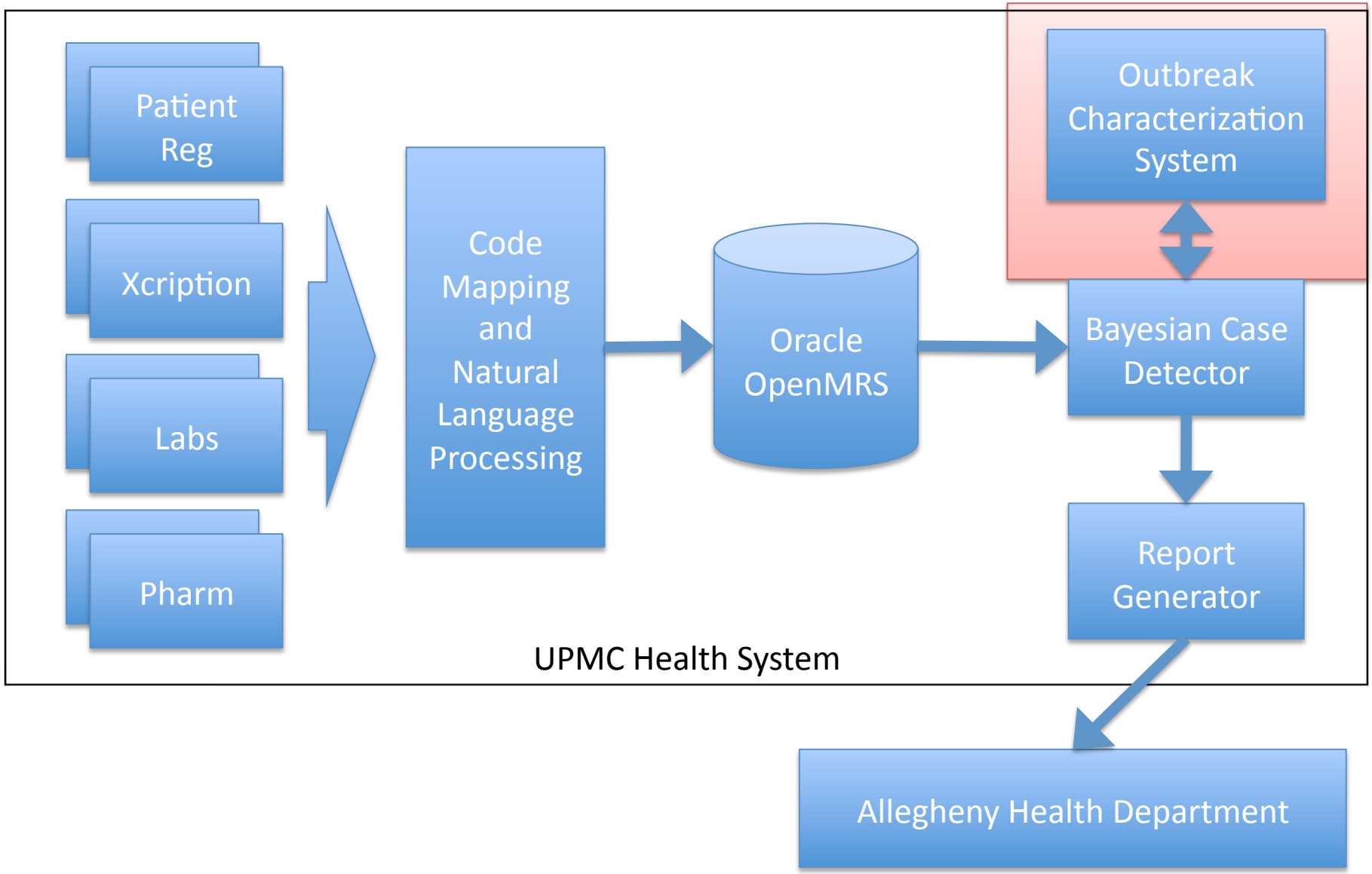
Our ILI Report was Provided to Media During Influenza Season



Probabilistic Case Notification

- Alert when Bayesian inference outputs a diagnosis with a probability over a threshold
  - The threshold depends on the disease!
- Use the posterior probabilities directly to create expected case counts
- Monitor expected case counts using time series analysis

Probabilistic Case Detection System



Outbreak Characterization System

- Project 2 of our Center of Excellence Proposal
- Utilize Bayesian methods to integrate the posterior probabilities of disease, environmental data and syndrome data to detect and characterize an outbreak
- Provides updated prior probabilities to the Bayesian Case Detector

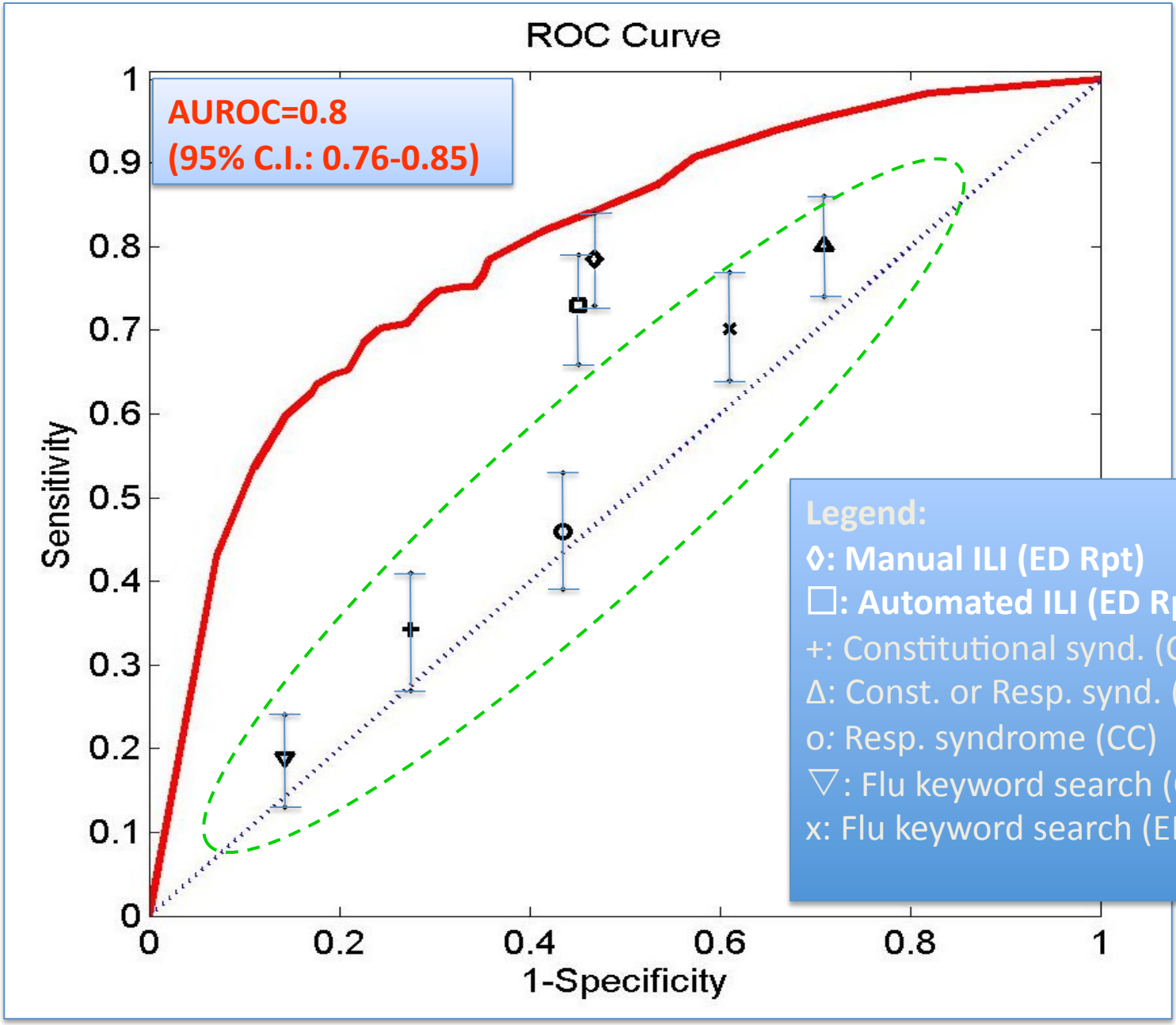
Pilot Evaluation:
Influenza Case Detection

Pilot Influenza Study

- Data: ED reports and chief complaints from 363 cases (181 positives and 182 negatives)
- Gold standard: laboratory confirmed reports
- Case Detection Methods
  - BCD + MedLEE
  - Keyword search (flu or influenza)
  - ILI classifier (using MedLEE+ILI rule)
  - Respiratory syndrome (using chief complaint only)
  - Constitutional syndrome (using chief complaint only)

Including measured temperature for fever

Results of Influenza Study



Comparison of Case Detection Methods

	ELR	Rule Based Case Detection	Probabilistic Case Detection
Data Source	Microbiology Lab	Entire Medical Record	Entire Medical Record
Use preliminary results / Orders	✖	✔	✔
Use clinical findings	✖	✔	✔
Identify probable cases	✖	✔	✔
Build case definitions from experts and/or data	✖	✔	✔
Incorporate disease prevalence (priors)	✖	✖	✔
Output case posterior probability	✖	✖	✔
Handle missing information	✖	✖	✔
Incorporate test sensitivity/ specificity	✖	✖	✔

Solution: Probabilistic Case Detection

- Use data from the entire patient visit including preliminary data
- Use structured and coded data; extracted when necessary from free text
- Utilize Bayesian diagnosis which integrates disease prevalence (prior probability) and findings to provide definitive as well as probable diagnoses

Our Team

- Rich Tsui
- Wendy Chapman
- Lee Christensen
- Mike Conway
- John Dowling
- Jeremy Espino
- Hendrik Harkema
- Qi Li
- Thomsun Sriburadej
- Howard Su
- Gregory Cooper
- Mike Wagner (PI)

Acknowledgements

- CDC funding: P01 HK000086 and 1U38 HK000063-01
- Dr. Ronald Voorhees at the Allegheny County Health Department